Parallel Imaging: A Unified Theoretical Framework for Image Generation*

WANG Kunfeng[1,2], LU Yue[1,3], WANG Yutong[1,3], WANG Fei-Yue[1,4]

1(The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190)

2(Parallel Vision Innovation Technology Center, Qingdao Academy of Intelligent Industries, Qingdao 266000)

3(School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 100049)

4(Research Center for Computational Experiments and Parallel Systems Technology, National University of Defense Technology, Changsha 410073)

***Abstract*—In order to build computer vision systems with good generalization capability, one usually needs large-scale, diversified labeled image data for learning and evaluating the in-hand computer vision models. Since it is difficult to obtain satisfactory image data from real scenes, in this paper we propose a unified theoretical framework for image generation, called parallel imaging. The core component of parallel imaging is software-defined artificial imaging systems. Artificial imaging systems receive small-scale image data collected from real scenes, and then generate large amounts of artificial image data. In this paper, we survey the realization methods of parallel imaging, including graphics rendering, image style transfer, generative models, and so on. Furthermore, we compare the properties of artificial images and actual images, and discuss the domain adaptation strategies.**

Keywords—Parallel Imaging, Model Learning, Graphics Rendering, Image Style Transfer, Generative Models

I. INTRODUCTION

Computer vision has developed rapidly in recent years, and it has been used in intelligent transportation, security monitoring, biometric authentication, commercial surveillance, human-machine interaction, and so on [1]–[4]. The core of computer vision is vision computing model (or visual model for short). To improve performance of computer vision system in practice, we need large-scale and diversified image data with labels to learn and evaluate the visual models [5]–[7]. However, collecting and labeling large-scale image data from real scenes is time-consuming, and only small-scale image data can be obtained and annotated, which fail to cover complex dynamic environments.

In 2016, Wang *et al.* extended the parallel system theory and ACP approach [8]–[14] to the vision computing field, and proposed the concept, framework and key techniques of parallel vision [15], [16]. Parallel vision focuses on constructing systematic theories and methods for visual perception and understanding of complex scenes, as shown in Figure 1. Artificial scenes are used to model and represent complex real scenes, and make it possible to collect and label large-scale and diversified image data. Computational experiments are utilized to train and evaluate a variety of visual models, and parallel execution is conducted to optimize the vision system online by virtual-real interaction. In parallel vision, the artificial virtual space is the new half space of solving complex vision problems. With real physical space, it composes the intact "complex space" of solving complex vision problems. Obtaining images meeting specific requirement from artificial virtual space is the first step of parallel vision.

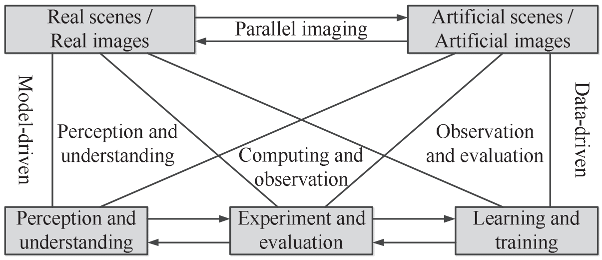

Fig. 1. Basic framework of parallel vision, for which parallel imaging provides real and artificial images

This paper proposes a unified framework of image generation, called parallel imaging. Parallel imaging is a branch of parallel vision, providing image data for parallel vision. The relationship between them is illustrated in Figure 2. The core component of parallel imaging is various software-defined artificial imaging systems. We obtain small-scale images from real scenes, and then put them into artificial imaging system to generate a large number of artificial images. These artificial and real images compose large-scale parallel images, which are then used to solve complex vision problem.

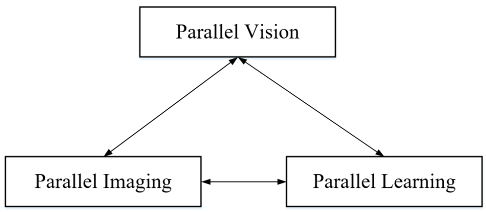

Fig. 2. Relationship between parallel imaging, parallel vision and parallel learning.

*Supported by National Natural Science Foundation of China (61533019, 71232006, 91520301)

The rest of this paper is organized as follows. Section 2 introduces the principles of parallel imaging. Section 3 summarizes the implementation methods, including graphics rendering, image style transfer and generative models. Section 4 analyzes the features of artificial and real images, and discusses domain adaptation. Section 5 provides an outlook of the developing trends of parallel imaging. Section 6 summarizes this paper.

II. FUNDAMENTAL OF PARALLEL IMAGING

Traditional image generation methods use cameras to take photos in real scenes, and then directly use these images or add random noise to the images as the input of the visual model. However, collecting and annotating large-scale image data from the real scene is not only time-consuming and laborious, but also difficult to guarantee the diversity of image data [17], [18]. In view of the limitations of traditional image generation methods, a parallel imaging theoretical framework is proposed in this paper.

Figure 3 shows the technological flowchart of the parallel imaging. Collect the real image "small data" from the real scene, input them to the artificial imaging system defined by the software, and automatically generate a large number of new artificial image data by analyzing and absorbing the characteristics of the real image. These artificial image data and the actual image "small data" together constitute the parallel image "big data" combining reality and virtuality. On the basis of computational experiments and parallel execution of parallel vision, get "little knowledge" used in some specific scenarios or tasks by learning. The "little" here means intelligent knowledge of the specific problems that need to be addressed, rather than a small amount of knowledge. The dashed box shows the core unit of the parallel imaging, i.e., the software-defined artificial image system.

III. METHOD FOR REALIZING PARALLEL IMAGING

There are many ways to realize the artificial image system: use computer graphics to synthesize images; build artificial scenes, and then render them; transfer the style of existing image to another to change its appearance; establish a generative model, directly output artificial images which meet the requirements. This section discusses three methods.

A. Graphics rendering

With the aid of advanced computer graphics, virtual reality, and micro simulation, we construct artificial scenes to simulate and represent complex and challenging actual scenes, and then render and generate artificial image with certain style [15], [16]. There have been many open source or commercial game engine and simulation tools, such as OpenStreetMap, CityEngine, and Unity3D, which can be used for artificial scene construction and graphics rendering.

Artificial scenes are made up of many elements, including static objects, moving objects, seasons, weather, light sources, and so on. For example, the elements of an artificial outdoor scene are shown in Table 1. Agent is used to represent the scene elements, and the multi-agent simulation is performed according to the physical rules. We use agent to represent objects and environments in an artificial scene. each agent has its own attributes.

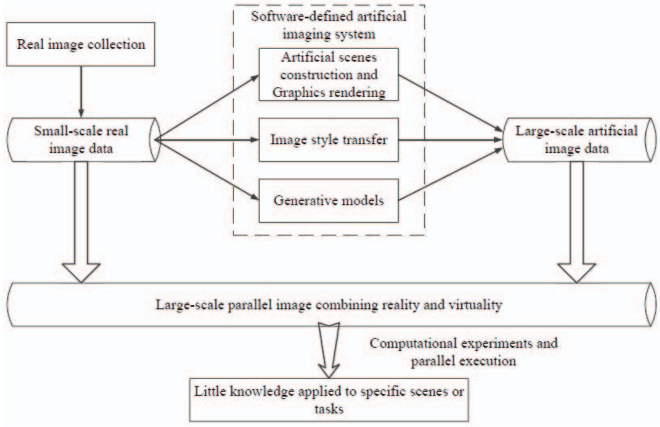

Fig. 3. Technological flowchart of parallel imaging.

TABLE I. COMPONENTS FOR ARTIFICIAL OUTDOOR SCENES

Scene element	content
Static object	buildings, sky, roads, sidewalks, fences, plants, columns, traffic signs, road markings, etc.
Dynamic object	cars (cars, vans, buses), bicycles, motorcycles, pedestrians, etc
Season	spring, summer, autumn and winter
Weather	overcast, rain, snow, fog, haze and so on
Light source	sun, street lamp, automobile lamp, etc.

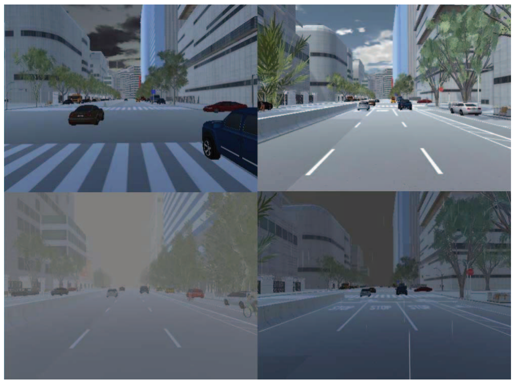

Fig. 4. Effects of graphics rendering given different weather conditions (Left top: cloudy weather; right top: sunny weather; left bottom: foggy weather; right bottom: raining weather.)

As shown in Figure 4, we set different weather conditions in artificial scene, render it and obtain artificial images. In addition, the actual scene is usually uncontrollable, non-repeatable, and difficult to annotate. Compared with it, the artificial scene is controllable, repeatable and automatic annotated. We can completely control the object appearance, motion pattern, illumination, weather condition, camera position, view angle and so on in the artificial scene, and generate various virtual images. Artificial scenes can also be used to accelerate the experiment. For example, for object detection and segmentation, we can quickly get diversified artificial images by changing the color of target object in consecutive frames, which is not possible in the real scene.

(a) Style images (b) Content images (c) New images

Fig. 5. Example effects of image style transfer.

B. Image style transfer

Image style transfer refers to the process of transferring the style of an image (style image) to another image (content image). The goal of style transfer is to obtain a new image, which is similar to the style image in artistic effect but consistent in content with content image [19]-[24]. With the style transfer technique, we can use existing images to generate lots of artificial images with different illumination, time, weather and season. In Figure 5, the first row shows the style image is a nocturnal city image (on the left), the content of image is a daytime city image (in the middle), style transferring can get a picture has the same illumination with the style image, and same scene layout with the content image (on the right).

At present, there are two main methods to realize image style transfer: texture synthesis and convolutional neural network. Other style transfer methods include image analogies [20] and color and texture transfer [21].

Efros *et al.* propose a texture synthesis method based on patch quilting [22]. This method partition content image into several patches, search the most matching patch in style image for each patch, and quilt them in an empty canvas. Ashikhmin *et al.* propose a method for speeding up this transfer [23].

Gatys *et al.* propose a style transfer method based on Convolutional Neural Networks (CNN) [24]. Firstly, a noisy image is deemed as the output image. Input the noise image, the style image and the content image respectively into the trained CNN, and get feature maps of each image. According to the loss function, compute the gradient of the output image using the error back-propagation, and iteratively update the output image.

Gatys *et al.* defined two loss functions: the error between the style image S and the output image O and the error between the content image I and the output image O. The final loss function is the weighted sum of both:

$$\mathcal{L}_{total} = \sum_{l=1}^{L} \alpha_l \mathcal{L}_c^l + \Gamma \sum_{l=1}^{L} \beta_l \mathcal{L}_s^l \quad (1)$$

$$\mathcal{L}_c^l = \frac{1}{2N_l D_l} \sum_{ij} \left(F_l[O] - F_l[I] \right)_{ij}^2 \quad (2)$$

$$\mathcal{L}_s^l = \frac{1}{2N_l^2} \sum_{ij} \left(G_l[O] - G_l[S] \right)_{ij}^2 \quad (3)$$

C. Generative models

The basic idea of generative models is to build a model of observation samples, and then use the model to predict. Generative adversarial networks (GANs) [25]-[32] are the advanced research of generative models with great ability of generating image.

Inspired by two-player zero-sum game, Goodfellow *et al.* [27] propose GANs in 2014. The computation procedure and structure of GAN is shown in Figure 6. Any differentiable function can be used as the generator and the discriminator. Here, we use differentiable functions D and G to represent the discriminator and the generator, and their inputs are real data x and random variables z, respectively. $G(z)$ represents the sample generated by G and obeying the distribution p_{data} of real data. If the input of discriminator D

is from the real data x, D should classify it to be true and label it as 1. If the input is from $G(z)$, D should classify it to be false and label it as 0. The purpose of D is to achieve correct classification of the data source: real or false, while the purpose of G is to make performance of the generated data $G(z)$ on D (i.e., $D(G(z))$ consistent with the performance of real data x on D (i.e., $D(x)$). The adversarial optimization process improves the performance of D and G gradually. Eventually, when the discrimination ability of D has been improved to a high level but cannot discriminate the data source correctly, it is thought that the generator G has captured the distribution of real data.

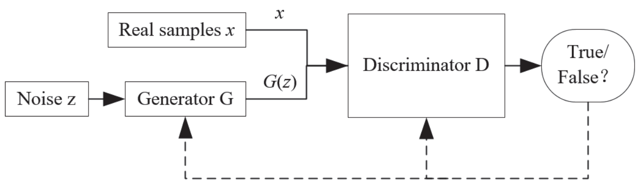

Fig. 6. Computation procedure and structure of GAN.

From the perspective of parallel imaging, GAN can be regarded as an artificial image generator, and artificial images can be obtained by image generation or translation. Parallel imaging applications are implemented primarily by the GAN derived model [28]–[32]. In this paper, we introduce three models closely related to parallel images.

Berthelot *et al.* [28] propose Boundary Equilibrium GAN (BEGAN). It draws on the strengths of both EBGAN [29] and WGAN [30], and achieves amazing effect under simple model and standard steps. BEGAN can generate high-quality human face with resolution of 128×128, including diverse gesture, expression, sex, color, light, beard and so on, as showed in Figure 7.

Shrivastava *et al.* [31] propose SimGAN to narrow the distribution gap between synthetic images and real images. They propose a simulated and unsupervised learning method that uses unlabeled real images to improve the fidelity of the synthesized images while maintaining the annotations of synthetic data. Figure 8 shows the structure of SimGAN, the synthesized image is processed by the Refiner neural network R to obtain a refined image, and discriminator D is used to discriminate whether the image is real or refined. Refiner network R and discriminator D update alternately.

Fig. 7. Face samples generated by BEGAN.

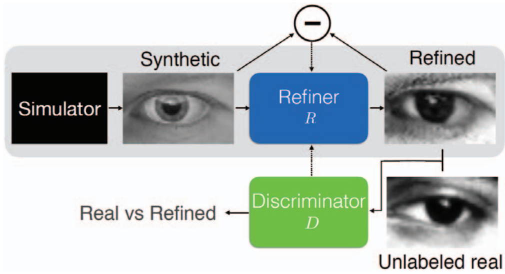

Fig. 8. Structure of SimGAN.

Zhu *et al.* [32] propose CycleGAN for learning to translate an image from a source domain to a target domain in the absence of paired examples. Duel to the loop consistency constraint in its structure, CycleGAN can transfer towards two directions without information lost, and generates realistic image with the original information obtained, avoiding mode collapse problem. CycleGAN is a general-purpose method and can be applied to a wide range of image-to-image translation tasks, including style transfer, object transfiguration, season transfer, and photo enhancement, as shown in Figure 9.

IV. Thoughts and Analyses

A. *Comparison of artificial images with real images*

Firstly, we hope artificial image can be more realistic. However, artificial images generated by graphical rendering have limitations on fidelity, and the value of a single artificial image is usually lower than that of a single real image. Luckily, GAN (e.g., CycleGAN) can translate an image from a source domain to a target domain, and can make the fidelity of artificial images infinity close to the real images which cannot be discriminated correctly by the discriminator.

Secondly, in diversity, artificial images have some advantages. The real world is diverse, and there are objects of various colors and movement pattern. However, because of the difficulty in data sample and annotation, the dataset of real scene is usually not various enough, especially lacking of data in bad lighting and weather conditions. Image style transfer, generative model and other methods can further improve the diversity of artificial images. Therefore, the artificial image dataset seems to be more diverse than the real image dataset.

Finally, the scale of artificial images is almost unlimited. The size of an artificial image can be arbitrarily large as long as the computing and storage resources are adequate. What's more important is that the true value annotation information of artificial images can be automatically generated without manual annotation that takes time and effort. Image style transfer, generative model and other methods can enlarge the scale of artificial images, which maintain or automatically obtain tag information for artificial images. Thus, the scale of a labeled artificial image dataset can be much larger than with a labeled real image dataset.

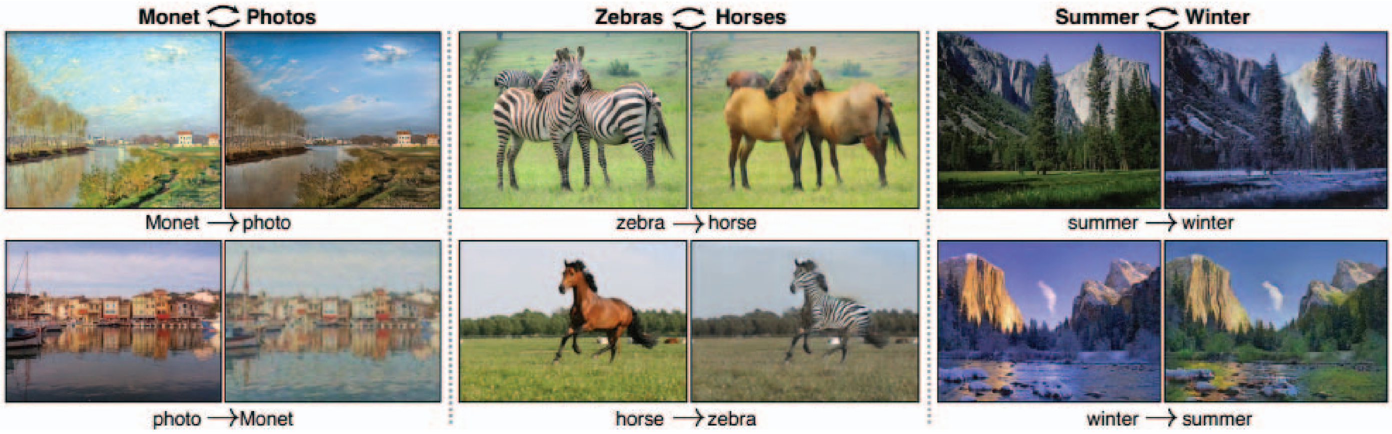

Fig. 9. Example results of CycleGAN for image-to-image translation.

In conclusion, artificial image systems have the potential to build large scale, diverse and labeled artificial image datasets to avoid overfitting and improve the generalization ability of vision model. Theoretically, the fidelity of artificial images can be infinitely close to that of the real image, and the discriminator can't even distinguish the two kinds of verisimilitude. Nevertheless, due to the fact that artificial images and real images come from different fields, there are dataset bias problems that require domain adaptation.

B. *Domain adaptation strategy*

Artificial image datasets can be used to learn and evaluate visual models. Because artificial images and real images come from different fields, there is a gap in the distribution (that is, Dataset bias). Three domain adaptation strategies can be used to handle dataset bias problems.

1) Unsupervised domain adaptation strategy: A large scale labeled artificial image data and large-scale unlabeled real image data are used to train the visual model. This strategy does not require any labeled data in the target domain.

2) Supervised domain adaptation strategy: A large scale labeled artificial image data and small-scale labeled real image data are used to train the visual model. This strategy doesn't use the unlabeled real image data.

3) Semi-supervised domain adaptation strategy: Large scale labeled artificial image data, small scale labeled real image data and large-scale unlabeled real image data are used to train the visual model. This strategy usually yields the best performance because it uses the most data.

V. Development Trend

We believe that the parallel imaging will be realistic, diverse, rapid and intelligent in the future, making the artificial image more vivid and diverse, faster generated, generation process more intelligent. Therefore, it is necessary to develop new mathematical tools and models to further strengthen the application of machine learning techniques such as deep learning, adversarial learning and parallel learning in image style transfer and generative model. Guide the direction of image generation by combine with the specific application scenarios, based on the idea of predictive learning and ensemble learning.

Traditionally, image generation has largely relied on sensing, but now it is increasingly dependent on computation. Modern image generation systems are closely integrated with sensing, optics, algorithms, computation, and so on. Emerging fields such as computational photography, computational microscopy, mobile and distributed imaging, together with algorithms such as compressive sensing and Bayesian inversion, have greatly expanded the scope of traditional image generation. Because of the deep integration of sensing and computing, a new interdisciplinary direction, Computational imaging, has been spawned. Computational imaging focuses not only on image processing, but also on image generation as well as the integration of sensing and computing. Since the publication of the International Journal IEEE Transactions on Computational Imaging[33] in 2015, the research and development of computational imaging have been remarkably accelerated.

The parallel imaging presented in this paper agrees well with the idea of computational imaging. Parallel imaging is a framework of image generation theory combining virtual and reality. It includes not only the real image generated by the sensor, but also the artificial image generated by the calculation. The artificial imaging system described in this paper has many methods of implementation, but it is similar to the computational imaging system as a whole. It is a combination of sensing, optics, algorithm and calculation. Therefore, the output of the imaging system can be calculated as an artificial image. We can draw on the research results in computational imaging field to deepen parallel imaging research.

In the hardware aspect, the real image is taken by the camera at present, and the artificial image is mainly generated on the computer. In the future, the degree of integration of image parallel hardware system will become increasingly high, and there will be a new camera, which can be called "parallel camera", with the real image and artificial image generated together, Parallel camera is embedded high-speed computing unit, integrated sensing (to generate real image) and calculating (to generate artificial image) function, which can output one channel of real images and multiple channels of artificial images at the same time, used for learning and evaluation of full of visual model directly.

VI. CONCLUSION

This paper presents a unified theoretical framework for image generation, called parallel imaging. As a branch of parallel vision, parallel imaging uses artificial images to expand the real image. This paper also summarizes the existing artificial image generation methods, including graphics rendering, image style transfer, generative models, and so on. The characteristics of the artificial image and the real image are compared and analyzed, and several domain adaptation strategies are discussed.

Parallel imaging theory is closely related to the current research hotspot of computational imaging. We predict that parallel imaging will develop in the direction of realism, diversity, efficiency, and intelligence in algorithms and software, and will develop in the direction of integration of sensing and computing in hardware. In the end, it is likely that a novel image generation apparatus, parallel camera that is able to capture one channel of real image and multiple channels of artificial images at the same time, will emerge in the near future.

References

[1] WANG K, LIU Y, GOU C, et al. A Multi-View Learning Approach to Foreground Detection for Traffic Surveillance Applications. IEEE Transactions on Vehicular Technology, 2016, 65(6): 4144−4158.

[2] WANG K, YAO Y. Video-Based Vehicle Detection Approach with Data-Driven Adaptive Neuro-Fuzzy Networks. International Journal of Pattern Recognition and Artificial Intelligence, 2015, 29(7), [32 pages], DOI: 10.1142/S0218001415550150.

[3] GOU C, WANG K, YAO Y, et al. Vehicle License Plate Recognition Based on Extremal Regions and Restricted Boltzmann Machines. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(4): 1096−1107.

[4] LIU Y, WANG K, SHEN D. Visual Tracking Based on Dynamic Coupled Conditional Random Field Model. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(3): 822−833.

[5] ImageNet. http://www.image-net.org/.

[6] The PASCAL Visual Object Classes Homepage. http://host.robots.ox.ac.uk/pascal/VOC/.

[7] COCO – Common Objects in Context. http://mscoco.org/.

[8] WANG F-Y. Parallel System Methods for Management and Control of Complex Systems. Control and Decision, 2004, 19(5): 485−489.

[9] WANG F-Y. Parallel Control and Management for Intelligent Transportation Systems: Concepts, Architectures, and Applications. IEEE Transactions on Intelligent Transportation Systems, 2010, 11(3): 630−638.

[10] WANG F-Y. Parallel Control: A Method for Data-Driven and Computational Control. Acta Automatica Sinica, 2013, 39(4): 293−302.

[11] MIAO Q, ZHU F, LV Y, et al. A Game-Engine-Based Platform for Modeling and Computing Artificial Transportation Systems. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(2): 343−353.

[12] WANG F-Y, ZHANG J J, ZHENG X, et al. Where Does AlphaGo Go: From Church-Turing Thesis to AlphaGo Thesis and Beyond. IEEE/CAA Journal of Automatica Sinica, 2016, 3(2): 113−120.

[13] ZHANG N, WANG F-Y, ZHU F, et al. DynaCAS: Computational Experiments and Decision Support for ITS. IEEE Intelligent Systems, 2008, 23(6): 19–23.

[14] WANG F-Y, WANG X, LI L, et al. Steps toward Parallel Intelligence. IEEE/CAA Journal of Automatica Sinica, 2016, 3(4): 345–348.

[15] WANG K, GOU C, WANG F-Y. Parallel Vision: An ACP-based Approach to Intelligent Vision Computing. Acta Automatica Sinica, 2016, 42(10): 1490–1500.

[16] WANG K, GOU C, ZHENG N, et al. Parallel Vision for Perception and Understanding of Complex Scenes: Methods, Framework, and Perspectives. Artificial Intelligence Review, vol. 48, no. 3, pp. 298−328, Oct. 2017.

[17] TORRALBA A, EFROS A A. Unbiased Look at Dataset Bias // 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, 2011: 1521−1528.

[18] MODEL I, SHAMIR L. Comparison of Data Set Bias in Object Recognition Benchmarks. IEEE Access, 2015, 3: 1953−1962.

[19] LUAN F, PARIS S, SHECHTMAN E, et al. Deep Photo Style Transfer. arXiv:1703.07511, 2017.

[20] HERTZMANN A, JACOBS C E, OLIVER N, et al. Image Analogies // Proc of the 28th annual conference on Computer graphics and interactive techniques, ACM, 2001: 327−340.

[21] OKURA F, VANHOEY K, BOUSSEAU A, et al. Unifying Color and Texture Transfer for Predictive Appearance Manipulation // Computer Graphics Forum (Proc of the Eurographics Symposium on Rendering), 2015, 34(4): 53−63.

[22] EFROS A A, FREEMAN W T. Image Quilting for Texture Synthesis and Transfer // Proc of the 28th annual conference on Computer graphics and interactive techniques, ACM, 2001: 341−346.

[23] ASHIKHMIN N. Fast Texture Transfer. IEEE Computer Graphics and Applications, 2003, 23(4): 38−43.

[24] GATYS L A, ECKER A S, BETHGE M. Image Style Transfer Using Convolutional Neural Networks // Proc of 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016: 2414−2423.

[25] WANG K, GOU C, DUAN Y, et al. Generative Adversarial Networks: The State of the Art and Beyond. Acta Automatica Sinica, 2017, 43(3): 321−332.

[26] WANG K, GOU C, DUAN Y, et al. Generative Adversarial Networks: Introduction and Outlook. IEEE/CAA Journal of Automatica Sinica, 2017, 4(4): 588−598.

[27] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative Adversarial Nets // Advances in Neural Information Processing Systems 27, Montreal, Quebec, Canada, 2014: 2672−2680.

[28] BERTHELOT D, SCHUMM T, METZ L. BEGAN: Boundary Equilibrium Generative Adversarial Networks. arXiv:1703.10717, 2017.

[29] ZHAO J, MATHIEU M, LECUN Y. Energy-based Generative Adversarial Network. arXiv:1609.03126, 2017.

[30] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein GAN. arXiv:1701.07875, 2017.

[31] SHRIVASTAVA A, PFISTER T, TUZEL O, et al. Learning from Simulated and Unsupervised Images through Adversarial Training. arXiv:1612.07828, 2016.

[32] ZHU J-Y, PARK T, ISOLA P, et al. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. arXiv:1703.10593, 2017.

[33] IEEE Transactions on Computational Imaging. http://signalprocessingsociety.org/publications-resources/ieee-transactions-computational-imaging.